日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００２年　５月１６日

出願番号
Application Number: 特願２００２－１４１３９０

[ST.10/C]: [ＪＰ２００２－１４１３９０]

出願人
Applicant(s): 科学技術振興事業団
株式会社国際電気通信基礎技術研究所

２００３年　４月　１日

特許庁長官
Commissioner,
Japan Patent Office 太田信一郎

出証番号　出証特２００３－３０２２８１６

【書類名】　特許願
【整理番号】　0020001
【特記事項】　特許法第３０条第１項の規定の適用を受けようとする特許出願
【提出日】　平成14年 5月16日
【あて先】　特許庁長官殿
【国際特許分類】　G10L 11/00
【発明者】
　【住所又は居所】　埼玉県川口市本町４丁目１番８号　科学技術振興事業団内
　【氏名】　パーハム　モクタリ
【発明者】
　【住所又は居所】　京都府相楽郡精華町光台二丁目２番地２　株式会社国際電気通信基礎技術研究所内
　【氏名】　ニック　キャンベル
【特許出願人】
　【識別番号】　396020800
　【氏名又は名称】　科学技術振興事業団
【特許出願人】
　【識別番号】　393031586
　【氏名又は名称】　株式会社国際電気通信基礎技術研究所
【代理人】
　【識別番号】　100099933
　【弁理士】
　【氏名又は名称】　清水　敏
【手数料の表示】
　【予納台帳番号】　173131
　【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【プルーフの要否】　要

出証特２００３－３０２２８１６

【書類名】　明細書

【発明の名称】　音声波形の特徴を高い信頼性で示す部分を決定するための装置およびプログラム、音声信号の特徴を高い信頼性で示す部分を決定するための装置およびプログラム、ならびに擬似音節核抽出装置およびプログラム

【特許請求の範囲】

【請求項１】　音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示す部分を決定するための装置であって、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、

前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発生源により良好に制御されている領域を推定するための推定手段と、

前記推定手段の出力と、前記発生源によって安定して発生されている領域として前記抽出手段により抽出され、かつ前記発生源によってその変化が良好に制御されていると前記推定手段によって推定された領域を前記音声波形の高信頼性部分として決定するための手段とを含む、音声波形の特徴を高い信頼性で示す部分を決定するための装置。

【請求項２】　前記抽出手段は、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で前記音声波形を節に分離するための手段と、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含む、請求項１に記載の装置。

【請求項３】　前記推定手段は、

出証特２００３－３０２２８１６

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、

前記データを用いて、前記線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第１の算出手段と、

前記線形予測手段の出力に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第２の算出手段と、

前記第１の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第２の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、請求項１に記載の装置。

【請求項４】　前記決定するための手段は、前記推定手段により前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出手段により抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するための手段を含む、請求項１～請求項３のいずれかに記載の装置。

【請求項５】　音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置であって、

前記音声信号の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で前記音声信号を擬似音節に分離するための手段と、

前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む、擬似音節核抽出装置。

【請求項６】　音声信号の特徴を高い信頼性で示す部分を決定するための装置であって、

前記音声信号に対する線形予測分析を行なうための線形予測手段と、

前記線形予測手段によるフォルマントの推定値と、前記音声信号とに基づいて

、前記フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第１の算出手段と、

前記線形予測手段による線形予測分析の結果に基づいて、前記音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第２の算出手段と

、

第１の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第２の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、音声信号の特徴を高い信頼性で示す部分を決定するための装置。

【請求項７】　音声波形のデータに基づいて、前記音声波形の特徴を高い信頼性で示す部分を決定するための装置としてコンピュータを動作させるプログラムであって、前記装置は、

前記データから前記音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および前記音声波形のピッチに基づいて、前記音声波形の各節のうち、前記音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、

前記データから前記音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、前記音声波形のうち、その変化が前記発生源により良好に制御されている領域を推定するための推定手段と、

前記推定手段の出力と、前記発生源によって安定して発生されている領域として前記抽出手段により抽出され、かつ前記発生源によってその変化が良好に制御されていると前記推定手段によって推定された領域を前記音声波形の高信頼性部分として決定するための手段とを含む、音声波形の特徴を高い信頼性で示す部分を決定するためのプログラム。

【請求項８】　前記抽出手段は、

前記データに基づいて、前記音声波形の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声波形の前記所定周波数領域のエネルギーの時間軸上の分布の波形の極

小部で前記音声波形を節に分離するための手段と、

前記音声波形のうち、各節内で、当該節内のエネルギーのピークを含み、かつ前記有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含む、請求項７に記載のプログラム。

【請求項９】　前記推定手段は、

前記音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、

前記データを用いて、前記線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第１の算出手段と、

前記線形予測手段の出力に基づいて、前記音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第２の算出手段と、

前記第１の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第２の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、請求項７に記載のプログラム。

【請求項１０】　前記決定するための手段は、前記推定手段により前記音声波形の変化が前記発生源により良好に制御されていると推定された領域のうち、前記抽出手段により抽出された領域に含まれる領域を前記音声波形の高信頼性部分として決定するための手段を含む、請求項７～請求項９のいずれかに記載のプログラム。

【請求項１１】　音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置としてコンピュータを動作させるプログラムであって、前記擬似音節核抽出装置は、

前記音声信号の各区間が有声区間か否かを判定するための有声判定手段と、

前記音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で前記音声信号を擬似音節に分離するための手段と、

前記音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ前記

有声判定手段により有声区間であると判定された区間であって、かつ前記所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む、擬似音節核抽出プログラム。

【請求項１２】　音声信号の特徴を高い信頼性で示す部分を決定するための装置としてコンピュータを動作させるプログラムであって、前記装置は、

前記音声信号に対する線形予測分析を行なうための線形予測手段と、

前記線形予測手段によるフォルマントの推定値と、前記音声信号とに基づいて、前記フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第１の算出手段と、

前記線形予測手段による線形予測分析の結果に基づいて、前記音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第２の算出手段と、

第１の算出手段により算出された前記フォルマント周波数の推定値の非信頼性の時間軸上の分布と、前記第２の算出手段により算出された前記音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、前記音声波形の変化が前記発生源により良好に制御されている領域を推定するための手段とを含む、音声信号の特徴を高い信頼性で示す部分を決定するためのプログラム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

この発明は、一般的には音声波形からその特徴を高い信頼性で示す部分を抽出するための技術に関し、特に、音声波形の発生源の状態を高い信頼性で推定するために有効な領域を、音声波形から抽出するための技術に関する。

【０００２】

【従来の技術】

［用語の定義１］

最初に、この節で使用される用語について定義する。

【０００３】

「緊張音」（ｐｒｅｓｓｅｄ　ｓｏｕｎｄ）とは、発声の際に声門が緊張して

いるために声門を気流が通過しにくく、かつ通過をする際の気流の加速度が大きくなるように発声される音のことをいう。この場合、声門気流波形はサインカーブから大きく変形し、その微分波形の傾きが局部的に大きくなる。音声がこうした特徴を有する場合、「緊張性」の音声であると呼ぶことにする。

【０００４】

「気息音」（ｂｒｅａｔｈｙ　ｓｏｕｎｄ）とは、発声の際に声門に緊張がないために気流が通過しやすく、その結果声門気流波形がサインカーブに近くなるように発声される音をいう。この場合、声門気流波形の微分波形の傾きが局部的に大きくなることはない。音声がこうした特徴を有する場合、「気息性」の音声であると呼ぶことにする。

【０００５】

「地声」（モーダル、ｍｏｄａｌ）とは、緊張音と気息音との中間の発声のことをいう。

【０００６】

「ＡＱ指数」（Ａｍｐｌｉｔｕｄｅ　Ｑｕｏｔｉｅｎｔ）とは、声門（声帯）気流の波形のピークツーピークの振幅を、声門気流の波形の微分の振幅の最小値で除した値のことをいう。

【０００７】

［従来の技術］

音声認識と並んで重要な音声研究分野に、音声合成がある。最近の信号処理技術の発達により、音声合成が既に多くの分野で利用されている。しかし、今までの音声合成は単にテキスト情報を音声化しているだけともいえ、人間が発話する際のような微妙な感情の表現までは行なえない。

【０００８】

たとえば、人間が発話する際には、怒り、喜び、および悲しみなどの情報が、発話内容以外の情報、つまり声色などにより伝達される。このように発話に付随する、言語以外の情報をパラ言語情報と呼ぶ。これらはテキスト情報のみでは表わせない情報である。しかし従来の音声合成では、こうしたパラ言語情報を伝達することは難しかった。マンマシンインタフェースをより効率的なものとするた

めには、テキスト情報だけではなくパラ言語情報も音声合成の際に伝えられるようにすることが望ましい。

【０００９】

こうした問題を解決するために、種々の発話スタイルで連続的に音声合成を行なおうとする試みがある。ひとつの具体的な方策として次のようなものがある。すなわち、発話を録音してデータ処理可能な形でデータベース化し、さらにその中で所望の特徴(怒り、喜び、悲しみなど)を表わすと思われる発話単位にそれらの特徴を示すラベルを付ける。音声合成の際には所望のパラ言語情報に対応したラベルが付けられた音声を利用する。

【００１０】

しかし、十分な広さの発話スタイルをカバーできるようにデータベースを構築しようとすれば、膨大な量の録音音声を処理しなければならない。そのために、自動的にオペレータの介在なく確実にそうした特徴の抽出とラベル付け処理とを行なえるようにする必要がある。

【００１１】

以下、パラ言語情報の一例を挙げる。発話スタイルの一つとして、緊張音と気息音という区別がある。緊張音では声門が緊張しているために、どちらかというと強い発声となる。一方気息音では、音声はサインカーブに近く、強いという印象はない。したがって緊張音と気息音という区別も発話スタイルの一つとして重要であり、その程度を数量化できれば、パラ言語情報として利用できる可能性がある。

【００１２】

緊張音と気息音との音質を区別する音響学的な指標については、今までにも数多くの研究がなされてきた。たとえば文末にリストした参考文献１を参照されたい。しかし、そうした研究の多くは、持続的に安定して母音を発音している間に録音された発話（または歌）を対象としたものに限定されていた。実際、膨大な量の発話の録音データから得られた音響測定データに基づいて、緊張性と気息性との程度を信頼性高く計量しなければならないというのは非常に大きな問題であり、かつ実現された場合には非常に有用となるであろう。

【００１３】

スペクトルドメインでの音源の属性を推定しようとする様々な手段が提案されて来たが、それよりも直接的な推定が、声門気流の波形とその導関数との組み合わせによって得られるはずである。そうした推定の一例が文末の参考文献２において提案されたＡＱ指数である。

【００１４】

参考文献２では、ＡＱ指数の一つの利点として、音圧レベル（ＳＰＬ）から比較的独立していること、およびその値が主として発音の質的なものに依存していることがあげられている。他の利点として考えられるのは、このパラメータが純粋に振幅ドメインのものであって、種々の発話スタイルに応じた、推定された声門波形の時間ドメインの特徴量を測定する際の誤差源に対して比較的免疫性があることである。また、参考文献２の著者らによれば、様々な発音スタイルで「ａ」という母音を持続して発音した場合、４人の男性と４人の女性との全てに対して、発音を気息性のものから緊張性に変えていくにしたがって、ＡＱ指数の値は単調に減少したとのことである（参考文献２の第１３６頁）。したがってＡＱ指数は、ここで我々が解決しようとしている問題に関して有効である可能性が高い。ただし、ＡＱ指数が有効となるためには、次の条件が満足される必要がある。

【００１５】

１）　録音された通常の発話について、ロバストでかつ信頼性高くＡＱ指数を測定できること、および

２）そうした条件で測定された知覚上の特徴が顕著な部分を確認することができること。

【００１６】

【発明が解決しようとする課題】

このような条件を満足させるためには、自然に発話された音声などの物理量を表わす音声波形から、いかにして信頼性高く音声波形の特徴を表わすパラメータを抽出できるかが重要である。特に音声の場合のように、発話が話者によりその細部まで完全にはコントロールされているわけでない場合、また様々な人が様々なスタイルで発話する場合には、パラメータを抽出すべき部分として信頼性がお

ける場所と、そうでない場所とが存在することが考えられる。そのため、音声波形のうちのどの部分を処理対象とするかが重要である。またそのために、日本語のように音節が発音の単位となる場合、音節の中心部（仮にこれを「音節核」と呼ぶ。）を誤りなく抽出できるようにすることが必要である。

【００１７】

したがって、本発明の目的は、音声波形の特徴を高い信頼性で示す部分を決定することを可能とすることである。本発明のほかの目的は、本発明のさらに他の目的は、音節核を高い信頼性で抽出できるようにすることである。

【００１８】

【課題を解決するための手段】

本発明の第１の局面は、複数個の節に分解可能な、物理的量を表わす音声波形のデータに基づいて、音声波形の特徴を高い信頼性で示す部分を決定するための装置と、そうした装置としてコンピュータを動作させるプログラムに関する。この装置は、データから音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および音声波形のピッチに基づいて、音声波形の各節のうち、音声波形の発生源によって安定して発生されている領域を抽出するための抽出手段と、データから音声波形のスペクトルの時間軸上の分布を算出し、当該スペクトルの時間軸上の分布に基づいて、音声波形のうち、その変化が発生源により良好に制御されている領域を推定するための推定手段と、推定手段の出力と、発生源によって安定して発生されている領域として抽出手段により抽出され、かつ発生源によってその変化が良好に制御されていると推定手段によって推定された領域を音声波形の高信頼性部分として決定するための手段とを含む。

【００１９】

抽出手段による抽出結果と、推定手段による推定結果との双方に基づいて音声波形の高信頼性部分を決定するので、決定結果がより確実なものとなる。

【００２０】

抽出手段は、データに基づいて、音声波形の各区間が有声区間か否かを判定するための有声判定手段と、音声波形の所定周波数領域のエネルギーの時間軸上の分布の波形の極小部で音声波形を節に分離するための手段と、音声波形のうち、

各節内で、当該節内のエネルギーのピークを含み、かつ有声判定手段により有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域を抽出するための手段とを含んでもよい。

【００２１】

有声と判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域が抽出されるので、発話者が安定して発声している区間を確実に抽出できる。

【００２２】

また好ましくは、推定手段は、音声波形に対する線形予測分析を行ないフォルマント周波数の推定値を出力するための線形予測手段と、データを用いて、線形予測手段によるフォルマント周波数の推定値の非信頼性の時間軸上の分布を算出するための第１の算出手段と、線形予測手段の出力に基づいて、音声波形の時間軸上のスペクトル変化の局所的な分散の、時間軸上の分布を算出するための第２の算出手段と、第１の算出手段により算出されたフォルマント周波数の推定値の非信頼性の時間軸上の分布と、第２の算出手段により算出された音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域を推定するための手段とを含む。

【００２３】

フォルマント周波数の推定値の非信頼性と、音声波形の時間軸上のスペクトル変化の局所的な分散との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域が推定される。振動変化の発生源（たとえば発話者）が明確な意図をもって振動を制御している領域が推定できるので、そうした領域から振動の特徴量を算出すれば、算出された特徴量の信頼性が高くなることが期待できる。

【００２４】

決定するための手段は、推定手段により音声波形の変化が発生源により良好に制御されていると推定された領域のうち、抽出手段により抽出された領域に含まれる領域を音声波形の高信頼性部分として決定するための手段を含んでもよい。

【００２５】

出証特２００３－３０２２８１６

音声波形の変化が発生源により良好に制御されていると推定された領域であって、かつ発生源により音声波形が安定に発生されているもののみを高信頼性部分として決定する。したがって真に信頼性が高い部分を抽出できる。

【００２６】

本発明の他の局面は、音声信号を擬似音節に分離し、さらに各擬似音節の核部分を抽出するための擬似音節核抽出装置と、そうした装置としてコンピュータを動作させるプログラムとに関する。この擬似音節核抽出装置は、音声信号の各区間が有声区間か否かを判定するための有声判定手段と、音声信号の所定周波数領域のエネルギーの時間的な分布の波形の極小部で音声信号を擬似音節に分離するための手段と、音声信号のうち、各擬似音節内でのエネルギーのピークを含み、かつ有声判定手段により有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域を当該擬似音節の核として抽出するための手段とを含む。

【００２７】

有声区間であると判定された区間であって、かつ所定周波数領域のエネルギーが所定のしきい値以上である領域が擬似音節の核として抽出されるので、発話者が安定して発声しているときの音声を抽出することができる。

【００２８】

本発明のさらに他の局面は、音声信号の特徴を高い信頼性で示す部分を決定するための装置と、そうした装置としてコンピュータを動作させるプログラムとに関する。当該装置は、音声信号に対する線形予測分析を行なうための線形予測手段と、線形予測手段によるフォルマントの推定値と、音声信号とに基づいて、フォルマントの推定値の非信頼性の時間軸上の分布を算出するための第１の算出手段と、線形予測手段による線形予測分析の結果に基づいて、音声信号のスペクトル変化の局所的な分散の時間軸上の分布を算出するための第２の算出手段と、第１の算出手段により算出されたフォルマント周波数の推定値の非信頼性の時間軸上の分布と、第２の算出手段により算出された音声波形のスペクトル変化の局所的な分散の時間軸上の分布との双方に基づいて、音声波形の変化が発生源により良好に制御されている領域を推定するための手段とを含む。

出証特２００３－３０２２８１６

【００２９】

フォルマントの推定値の非信頼性の時間軸上の分布も、音声信号のスペクトル変化の局所的な分散の時間軸上の分布も、その極小部ではいずれも音声信号のうちでその発生源により音声波形の発生が良好に制御されている部分を示す。これらの双方を用いて領域を推定するので、音声波形の発生が良好に制御されている部分を信頼性高く特定することができる。

【００３０】

【発明の実施の形態】

以下に述べる本発明の実施の形態は、コンピュータおよびコンピュータ上で動作するソフトウェアにより実現される。もちろん、以下に述べる機能の一部又は全部を、ソフトウェアでなくハードウェアで実現することも可能である。

【００３１】

［用語の定義２］

以下、本実施の形態の説明で使用される用語について定義する。

【００３２】

「擬似音節」とは、音声信号から所定の信号処理によって決定される信号の切れ目のことを指し、日本語音声の場合の音節を推定したものに対応する。

【００３３】

「ソノラントエネルギー」とは、音声信号のうちで、所定周波数（たとえば６０Ｈｚ～３ｋＨｚの周波数領域）のエネルギーのことをいい、デシベルで表わされる。

【００３４】

「信頼性の中心」（ｃｅｎｔｅｒ　ｏｆ　ｒｅｌｉａｂｉｌｉｔｙ）とは、音声波形に対する信号処理の結果、音声波形のうちで、対象となる音声波形の特徴を信頼性高く抽出することができるとみなされることとなった領域のことをいう。

【００３５】

「ディップ」とは、グラフなどの図形がくぼれた部分のことをいう。特に、時間の関数として変化するような値の時間軸上の分布により形成される波形のうち

、極小値に対応する部分をいう。

【００３６】

「非信頼性」とは、信頼性のなさを表わす尺度のことをいう。非信頼性は信頼性の逆の概念である。

【００３７】

図１に、本実施の形態で利用されるコンピュータシステム２０の外観図を、図２にコンピュータシステム２０のブロック図を、それぞれ示す。なおここに示すコンピュータシステム２０はあくまで一例であり、この他にも種々の構成が可能である。

【００３８】

図１を参照して、コンピュータシステム２０は、コンピュータ４０と、いずれもこのコンピュータ４０に接続されたモニタ４２、キーボード４６、およびマウス４８を含む。コンピュータ４０にはさらに、ＣＤ－ＲＯＭ（Ｃｏｍｐａｃｔ Ｄｉｓｃ Ｒｅａｄ－Ｏｎｌｙ Ｍｅｍｏｒｙ）ドライブ５０と、ＦＤ（Ｆｌｅｘｉｂｌｅ Ｄｉｓｋ）ドライブ５２とが内蔵されている。

【００３９】

図２を参照して、コンピュータシステム２０はさらに、コンピュータ４０に接続されるプリンタ４４を含むが、これは図１には示していない。またコンピュータ４０はさらに、ＣＤ－ＲＯＭドライブ５０およびＦＤドライブ５２に接続されたバス６６と、いずれもバス６６に接続された中央演算装置（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ：ＣＰＵ）５６、コンピュータ４０のブートアッププログラムなどを記憶したＲＯＭ（Ｒｅａｄ－Ｏｎｌｙ Ｍｅｍｏｒｙ）５８、ＣＰＵ５６が使用する作業エリアおよびＣＰＵ５６により実行されるプログラムの格納エリアを提供するＲＡＭ(Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）６０、および後述する音声データベースを格納したハードディスク５４を含む。

【００４０】

以下に述べる実施の形態のシステムを実現するソフトウェアは、たとえば、ＣＤ－ＲＯＭ６２のような記録媒体上に記録されて流通し、ＣＤ－ＲＯＭドライブ

５０のような読取装置を介してコンピュータ４０に読込まれ、ハードディスク５４に格納される。ＣＰＵ５６がこのプログラムを実行する際には、ハードディスク５４からこのプログラムを読み出してＲＡＭ６０に格納し、図示しないプログラムカウンタによって指定されるアドレスから命令を読出して実行する。ＣＰＵ５６は、処理対象のデータをハードディスク５４から読出し、処理結果を同じくハードディスク５４に格納する。

【００４１】

コンピュータシステム２０の動作自体は周知であるので、ここではその詳細については繰り返さない。

【００４２】

なお、ソフトウェアの流通形態は上記したように記憶媒体に固定された形には限定されない。たとえば、ネットワークを通じて接続された他のコンピュータからデータを受取る形で流通することもあり得る。また、ソフトウェアの一部が予めハードディスク５４中に格納されており、ソフトウェアの残りの部分をネットワーク経由でハードディスク５４に取込んで実行時に統合するような形の流通形態もあり得る。

【００４３】

一般的に、現代のプログラムはコンピュータのオペレーティングシステム（ＯＳ）によって提供される汎用の機能を利用し、それらを所望の目的にしたがって組織化した形態で実行することにより前記した所望の目的を達成する。したがって、以下に述べる本実施の形態の各機能のうち、ＯＳまたはサードパーティが提供する汎用的な機能を含まず、それら汎用的な機能の実行順序の組合せだけを指定するプログラム（群）であっても、それらを利用して全体的として所望の目的を達成する制御構造を有するプログラム(群）である限り、それらが本発明の技術的範囲に含まれることは明らかである。

【００４４】

本実施の形態のプログラムを装置とみなして機能的に示したのが図３以下のブロック図である。図３を参照して、この装置８０は、ハードディスク５４に格納された音声データ８２に対して以下に説明する処理を行なって、音声データに含

まれる各処理単位（たとえば音節）ごとに前述したＡＱ指数を算出し出力するためのものである。なお、音声データは後述するように１フレーム３２ｍｓｅｃとなるように予めフレーム化されている。

【００４５】

装置８０は、音声データに対して高速フーリエ変換（Ｆａｓｔ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ：ＦＦＴ）を行なうＦＦＴ処理部９０と、ＦＦＴ処理部９０の出力を用い、音声データにより表わされる音声波形のうちの６０Ｈｚ～３ｋＨｚの周波数領域のエネルギーの時間的変化および音声のピッチの変化に基づいて、音声データにより表わされる音声波形の各音節のうち、話者の発声機構によって安定して発生されている領域（これを以後「擬似音節核」と呼ぶ。）とを抽出する音響・韻律分析部９２と、音声データ８２に対してケプストラム分析を行ない、さらに、ＦＦＴ処理部９０の出力を用いてケプストラム分析の結果音声スペクトルの変化が少なく、音声データの特徴を信頼性高く抽出できると思われる部分（これを「高信頼性・小変動部の中心」または「高信頼・小変動の中心」または単に「信頼性の中心」と呼ぶ。）を推定するためのケプストラム分析部９４とを含む。

【００４６】

装置８０はさらに、ケプストラム分析部９４の出力する信頼性の中心（高信頼性・小変動部の中心）の中で、音響・韻律分析部９２の出力する擬似音節核の中にあるものだけを擬似音節中心として抽出するための擬似音節中心の抽出部９６と、擬似音節中心の抽出部９６によって抽出された擬似音節中心に対応する音声データに対して、フォルマントの初期推定と最適化処理とを行なって最終的なフォルマントの推定値を出力するためのフォルマントの最適化部９８と、音声データに対して、フォルマントの最適化部９８から出力されるフォルマント値を用いた適応的フィルタ処理などの信号処理を行なって声門気流波形の微分を推定し、さらにそれを積分することによって声門気流波形を推定し、それらに基づいてＡＱ指数を計算するためのＡＱ指数計算部１００とを含む。

【００４７】

図４は、音声データの構成を模式的に示す図である。図４を参照して、音声デ

ータ波形１０２は、それぞれ３２ｍｓｅｃごとのフレームに分けられ、かつ前後のフレーム間では８ｍｓｅｃごとにずらしてデジタル化されている。そして、後述する処理では、たとえばある時点ｔ０では第１のフレームを先頭として処理をし、次の時点ｔ１では８ｍｓｅｃずれた次の第２のフレームを先頭として処理をする、という形で処理を行なっていく。

【００４８】

図５は、図３に示す音響・韻律分析部９２のブロック図である。図５を参照して、音響・韻律分析部９２は、音声波形から測定される音源のピッチを用いて、処理対象のフレームが有声区間か否かを判定する（この方法については参考文献３を参照）ためのピッチ判定部１１０と、ＦＦＴ処理部９０の出力に基づいて所定周波数領域（６０Ｈｚ～３ｋＨｚ）のソノラントエネルギの時間軸上の波形分布を算出するためのソノラントエネルギ算出部１１２と、ソノラントエネルギ算出部１１２によって算出されるソノラントエネルギの時間軸上の分布波形の輪郭に対して凸包アルゴリズムを適用することにより、ソノラントエネルギの時間軸上の分布波形の輪郭の中のディップを検出して、入力音声を擬似音節に分割する（この方法については参考文献４および５を参照）ためのディップ検出部１１４と、ディップ検出部１１４によって得られた擬似音節中の、ソノラントエネルギの最大値（ＳＥpeak）が得られる点を起点として、その左右に、ソノラントエネルギが所定のしきい値（０．８×ＳＥpeak）より大きく、かつピッチ判定部１１０によって有声区間であると判定されたフレームであって、かつ同じ擬似音節中のフレームを１フレームずつ広げていくことにより、擬似音節核を出力するための有声・エネルギ判定部１１６とを含む。

【００４９】

図６は、図３に示すケプストラム分析部９４のブロック図である。図６を参照して、ケプストラム分析部９４は、音声データ８２の音声波形に対して選択的線形予測（Ｓｅｌｅｃｔｉｖｅ　Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｏｎ：ＳＬＰ）分析を行なって、ＳＬＰケプストラム係数$c_{f,i}$を出力するための線形予測分析部１３０と、このケプストラム係数に基づいて先頭の４つのフォルマントの周波数と帯域との初期推定値を算出するためのフォルマント推定部１３２とを含む。

フォルマント推定部１３２は、参考文献６により提案された線形ケプストラムーフォルマントマッピングを利用し、かつ同じデータのサブセットを使用して注意深く測定された母音フォルマントに対するマッピングを学習させてある。この学習については、参考文献７を参照されたい。

【００５０】

ケプストラム分析部９４はさらに、推定されたフォルマント周波数などに基づいてケプストラム係数C_i^{simp}を再計算するためのケプストラム再生成部１３６と、ＦＦＴ処理部９０の出力に対して対数変換およびコサイン逆変換（ＩＤＣＴ）を行なってＦＦＴケプストラム係数を算出するための対数変換および逆ＤＣＴ部１４０と、ケプストラム再生成部１３６により計算されたケプストラム係数C_i^{simp}と、対数変換および逆ＤＣＴ部１４０により計算されたＦＦＴケプストラム係数C_i^{FFT}との間の差を表わす値として次の式により定義されるケプストラム距離d_f^2を計算し、フォルマント推定部１３２によって推定されたフォルマント周波数などの値の非信頼性を表わす指標として出力するためのケプストラム距離計算部１４２とをさらに含む。

【００５１】

【数１】

$$d_f^2 = \mathrm{Sum}_i\{i^2.(c_i^{simp} - c_i^{FFT})^2\}$$

フォルマント推定部１３２、ケプストラム再生成部１３６、ケプストラム距離計算部１４２、および対数変換および逆ＤＣＴ部１４０により、線形予測分析の結果に基づいて推定されたフォルマント周波数などの値の非信頼性が算出される。

【００５２】

ケプストラム分析部９４はさらに、線形予測分析部１３０の出力するケプストラム係数からΔケプストラムを算出する為のΔケプストラム算出部１３４と、Δケプストラム算出部１３４の出力するΔケプストラムに基づいて、各フレームごとに、そのフレームを含む５フレームのスペクトル変化の大きさの分散を算出する為のフレーム間分散算出部１３８とを含む。フレーム間分散算出部１３８の出

力は、局所的なスペクトルの動きの時間軸上の分布波形の輪郭を表わすものとなり、その極小値は、参考文献８で提案されている調音音声学理論にならっていえば、制御された動きＣＭ（Controlled Movement）を示すものと考えることができる。

【００５３】

さらにケプストラム分析部９４は、ケプストラム距離計算部１４２の出力するフォルマント周波数の推定値の非信頼性を示す値と、フレーム間分散算出部１３８の出力する各フレームごとの局所的なフレーム間分散値とを受け、両者の値を規格化し統合して、フレームごとの音声信号の非信頼性を示す値の時間軸上の分布波形として出力するための規格化および統合部１４４と、規格化および統合部１４４の出力する非信頼性の値の時間軸上の分布波形により形成される波形の輪郭のディップを凸包アルゴリズムにより検出して、信頼性の中心候補として出力するための信頼性の中心候補出力部１４６とを含む。

【００５４】

図７は、図６に示す規格化および統合部１４４のブロック図である。図７を参照して、規格化および統合部１４４は、ケプストラム距離計算部１４２により出力されたケプストラム距離を［０，１］の値に規格化するための第１の規格化部１６０と、フレーム間分散算出部１３８が各フレームごとに算出するフレーム間分散の値を［０，１］の値に規格化するための第２の規格化部１６２と、局所的なフレーム間分散の値の時間軸上の位置を、ケプストラム距離計算部１４２の出力するケプストラム距離のサンプリングタイミングと一致させるように線形補間処理を行なうための補間処理部１６４と、第１の規格化部１６０の出力と補間処理部１６４の出力とを１フレームごとに平均して出力するための平均計算部１６６とを含む。平均計算部１６６の出力は、統合された値の時間軸上の分布波形の輪郭を表わす。信頼性の中心候補出力部１４６によってこの波形の輪郭のディップ（極小部）を検出することにより、非信頼性が最も低い部分（信頼性が最も高い部分）を信頼性の中心の候補として特定することができる。

【００５５】

図８は、図３に示すフォルマントの最適化部９８のブロック図である。図８を

参照して、フォルマントの最適化部９８は、音声波形に対してＦＦＴ処理を行なうためのＦＦＴ処理部１８０と、ＦＦＴ処理部１８０の出力に対して対数変換およびコサイン逆変換を行なうための対数変換および逆ＤＣＴ部１８２と、対数変換および逆ＤＣＴ部１８２の出力するＦＦＴケプストラム係数と、後述するフォルマントの推定値との間の距離を計算するためのケプストラム距離計算部１８４と、信頼性の中心候補の各々における第１～第４のフォルマント周波数の初期推定値を初期値とし、ケプストラム距離計算部１８４が計算する距離を最小にするように山登り法によってフォルマントの推定値を最適化するための距離最小化処理部１８６とを含む。距離最小化処理部１８６によって最適化されたフォルマント推定値がフォルマントの最適化部９８の出力としてＡＱ指数計算部１００に与えられる。

【００５６】

図９を参照して、ＡＱ指数計算部１００は、音声信号のうちで音節中心に相当する位置の６４ｍｓｅｃの部分のうち、７０Ｈｚ以上の周波数成分のみを選択的に通過させるためのハイパスフィルタ２００と、ハイパスフィルタ２００の出力のうち、最適化された第４フォルマント周波数とその帯域との和以下の周波数成分のみを選択的に通過させるための適応的ローパスフィルタ２０２と、適応的ローパスフィルタ２０２の出力に対し、第１～第４フォルマント周波数を用いた適応的逆フィルタ処理を行なうための適応的逆フィルタ２０４とを含む。適応的逆フィルタ２０４の出力は、声門気流波形の微分波形となる。

【００５７】

ＡＱ指数計算部１００はさらに、適応的逆フィルタ２０４の出力を積分して声門気流波形を出力するための積分回路２０６と、積分回路２０６の出力のピークツーピークの最大振幅を検出するための最大ピーク間振幅検出回路２０８と、適応的逆フィルタ２０４の出力の負のピークの最大振幅を検出するための最大の負のピーク振幅検出回路２１０と、最大の負のピーク振幅検出回路２１０の出力に対する最大ピーク間振幅検出回路２０８の出力の比を算出するための比計算回路２１２とを含む。比計算回路２１２の出力がＡＱ指数である。

【００５８】

図１～図９に示した装置は以下のように動作する。まず、使用された音声データ８２について説明する。この音声データは参考文献９で使用されたものであり、日本語のネイティブスピーカである女性が３つの物語を読んだものを録音して作成されたものである。この物語は、怒りと、喜びと、悲しみという感情を引き起こすように予め作成されていたものである。物語の各々は４００文の長さ（およそ３０，０００音素）以上の発話を含む。各発話は別々の音声波形ファイルに格納され処理された。

【００５９】

各文の発話データはＦＦＴ処理部９０によるＦＦＴ処理の後、以下のようにして処理される。処理は大きく見て二つの系統に分かれ実行される。第１の系統は音響・韻律分析部９２で行なわれる音響韻律的な処理であり、他の系統はケプストラム分析部９４が行なう音響音声学的な処理である。

【００６０】

音響韻律的な系統の処理では、図５に示すソノラントエネルギ算出部１１２によって６０Ｈｚ～３ｋＨｚ周波数領域のソノラントエネルギが算出される。ソノラントエネルギ算出部１１２の出力する一文の発話データの全体波形の輪郭から、ディップ検出部１１４が凸包アルゴリズムによりディップを検出する。このディップにより、この発話文は擬似音節に分割される。

【００６１】

有声・エネルギ判定部１１６は、擬似音節の中でソノラントエネルギが最大（ＳＥpeak）となる点を見つける。この点が擬似音節核の初期点である。有声・エネルギ判定部１１６はさらに、この擬似音節核の初期点から始めて、その左右に向かい、ソノラントエネルギが０．８×ＳＥpeak以下のフレーム、またはピッチ判定部１１０が有声でないと判定したフレーム、または擬似音節の外のフレームに出会うまで、擬似音節核の範囲を広げる。こうして擬似音節核の境界が決定される。この情報は擬似音節中心の抽出部９６に与えられる。なお、ここでしきい値として０．８の値を用いているが、これは単なる例であって，応用によりこのしきい値を適切な値に代える必要がある。

【００６２】

図６を参照して、入力された一つの発話文に対して線形予測分析部１３０が線形予測分析を行ない、ＳＬＰケプストラム係数を出力する。Δケプストラム算出部１３４がこのＳＬＰケプストラム係数に基づいてΔケプストラムを算出し、フレーム間分散算出部１３８に与える。フレーム間分散算出部１３８は、このΔケプストラム係数に基づき、各フレームごとに、そのフレームを含む５フレームの中での局所的なスペクトル変化の分散を計算する。この分散が小さいほど発話者の発声が発話者によりよく制御されていると考えられ、逆にこの分散が大きいと話者による制御がよくされていないと考えられるので、フレーム間分散算出部１３８の出力は発話者の発声が信頼されない程度（非信頼性）を表わすと考えられる。

【００６３】

図６をさらに参照して、フォルマント推定部１３２は、線形ケプストラムフォルマントマッピングを用い、ＳＬＰケプストラム係数に基づいて第１～第４フォルマントの周波数と帯域とを推定する。ケプストラム再生成部１３６は、フォルマント推定部１３２により推定された第１～第４フォルマントに基づいて逆にケプストラム係数を算出しケプストラム距離計算部１４２に与える。対数変換および逆ＤＣＴ部１４０は、フォルマント推定部１３２およびケプストラム再生成部１３６が処理したのと同じフレームのもとの音声データに対して対数変換およびコサイン逆変換を行なってＦＦＴケプストラム係数を算出しケプストラム距離計算部１４２に与える。ケプストラム距離計算部１４２は、ケプストラム再生成部１３６からのケプストラム係数と対数変換および逆ＤＣＴ部１４０からのケプストラム係数との間の距離を前述の「数１」の式にしたがって計算する。この結果得られるのは、フォルマント推定部１３２が推定したフォルマントの非信頼性を示す値の時間軸上の分布を表わす波形と考えられる。ケプストラム距離計算部１４２は、この結果を規格化および統合部１４４に与える。

【００６４】

図７を参照して、規格化および統合部１４４の第１の規格化部１６０は、図６のケプストラム距離計算部１４２の出力する、フォルマントの推定値から算出された各フレームごとの非信頼性値を［０，１］の範囲に正規化して平均計算部１

６６に与える。第２の規格化部１６２は、図６のフレーム間分散算出部１３８が出力する、フレームごとに計算された局所的なフレーム間分散の値を［０，１］の範囲に正規化して補間処理部１６４に与える。補間処理部１６４は、第２の規格化部１６２の各値に対し、第１の規格化部１６０の出力する各フレームのサンプリングポイントに対応する値が得られるように線形補間処理を行なって平均計算部１６６に与える。平均計算部１６６は、フレームごとに、第１の規格化部１６０の出力と補間処理部１６４の出力とを正規化し、その結果を時間軸上の非信頼性の分布を示す統合された波形として信頼性の中心候補出力部１４６に出力する。

【００６５】

信頼性の中心候補出力部１４６は、凸包アルゴリズムにより、規格化および統合部１４４の出力する統合された波形の輪郭のディップを検出して、そのフレームを特定する情報を図３の擬似音節中心の抽出部９６に対して信頼性の中心の候補として出力する。

【００６６】

図３に示す擬似音節中心の抽出部９６は、図６に示す信頼性の中心候補出力部１４６から与えられた信頼性の中心の中で、音響・韻律分析部９２から与えられた擬似音節核の中にあるもののみを擬似音節中心として抽出する。

【００６７】

以上の処理によって、音声データのうちで音声データの特徴を抽出する、または音声データをラベル付けするために適した高信頼性・小変動領域を示す情報が得られたことになる。したがって、この情報によって特定されるフレームについて所望の処理を行なえばよい。本実施の形態の装置では、擬似音節中心の抽出部９６はこの情報をフォルマントの最適化部９８に与え、フォルマントの最適化部９８はこの情報を用いて、以下のようにして擬似音節中心におけるＡＱ指数を算出する。

【００６８】

なお、本実施の形態の装置では、擬似音節中心の長さは連続する５フレームとする。１フレームは３２ｍｓｅｃであり、連続するフレームは互いに８ｍｓｅｃ

ずつずれているから、５フレームの全体では６４ｍｓｅｃの音声期間に相当する。

【００６９】

これらの擬似音節中心におけるＡＱ指数は、図９のＡＱ指数計算部１００中で得られる声門気流の波形により直接計算することができる。しかし、声門気流の推定自体、もともとのフォルマントに相当する声道の共振によって影響されており、その信頼性は共振の影響をもとの音声波形の６４ｍｓｅｃのデータから取り除くことができるかに依存している。したがって、そのような計算によって得られたＡＱ指数は信頼できないものとなる。

【００７０】

一方、擬似音節中心におけるフォルマントは、スペクトルがよく一致しているという意味で、既によい推定となっているが、本実施の形態の装置では、さらに以下のようにしてフォルマント周波数を最適化する。

【００７１】

すなわち、図８を参照して、ＦＦＴ処理部１８０は音声波形に対してフレームごとにＦＦＴ処理を行なう。対数変換および逆ＤＣＴ部１８２はＦＦＴ処理部１８０の出力に対して対数変換およびコサイン逆変換を行なう。ケプストラム距離計算部１８４は、対数変換および逆ＤＣＴ部１８２の出力するケプストラム係数と距離最小化処理部１８６から与えられるケプストラム係数の推定値との間の距離を計算する。距離最小化処理部１８６は、フォルマントの推定値を表わすケプストラム係数の値を起点として、ケプストラム距離計算部１８４により計算される距離が最小値となるように山登り法によって距離最小化処理部１８６から与えられたケプストラム係数の値をさらに最適化し、最小値が得られるときのフォルマント推定値を出力する。

【００７２】

ＡＱ指数計算部１００の内部構成は図９に示されており、この図９を参照して、擬似音節中心における音声データはまずハイパスフィルタ２００を通り、その結果７０Ｈｚ以下の低周波数の雑音が除去される。さらに適応的ローパスフィルタ２０２によって第４フォルマントより高い周波数領域のスペクトル情報が除去

される。そして、適応的逆フィルタ２０４によって第１～第４フォルマントによる影響が除去される。

【００７３】

その結果、適応的逆フィルタ２０４の出力は声門気流の波形の微分のよい推定値となる。これを積分回路２０６で積分することにより声門気流の波形の推定値が得られる。最大ピーク間振幅検出回路２０８によって声門気流の波形のピークツーピークの振幅の最大値を検出する。最大の負のピーク振幅検出回路２１０によって声門気流の微分波形のサイクル内での負の最大の振幅を検出する。最大ピーク間振幅検出回路２０８の出力の、最大の負のピーク振幅検出回路２１０の出力に対する比を比計算回路２１２で計算することにより、擬似音節中心におけるＡＱ指数が得られる。

【００７４】

こうして得られたＡＱ指数は、各擬似音節中心におけるもとの音声データの特徴（緊張音一気息音の間の度合い）を信頼性高くあらわしている。これら各擬似音節中心に対してＡＱ指数を計算し、さらにこれら得られたＡＱ指数を補間することにより、擬似音節中心以外の部分のＡＱ指数を推定することもできる。そうすることにより、音声データのうち、一定のＡＱ指数を示す部分に、当該ＡＱ指数に対応した適切なラベルをパラ言語情報として付けておき、音声合成の際には、所望のＡＱ指数を有する音声データを使用すれば、単なるテキストだけでなく、パラ言語情報をも含んだ形での音声合成を行なうことが可能になる。

【００７５】

図１０～図１２に、本実施の形態の装置をコンピュータにより実現した際の画面表示例を示す。

【００７６】

図１０を参照して、このプログラムによる表示ウィンドウには、音声データ波形２４０と、音声データに対して付された音声ラベル２４２と、基本周波数の波形の時間軸上の分布波形の輪郭２４４と、ソノラントエネルギの変動の時間軸上の分布波形の輪郭２４６と、Δケプストラムから計算されたスペクトル変化の局所的な分散の時間軸上の分布波形の輪郭２４８と、フォルマントーＦＦＴケプス

トラム距離の時間軸上の分布波形の輪郭２５０と、スペクトル変化の局所的な分散の分布波形の輪郭２４８およびフォルマントーケプストラム距離の分布波形の輪郭２５０を統合した波形である非信頼性の時間軸上の分布波形の輪郭２５２と、上述のようにして算出された擬似音節中心での声門のＡＱ指数２５４と、各擬似音節中心で推定された声道の面積関数２５６とが示されている。

【００７７】

音声データ波形２４０の表示領域に示された太い縦線２３２と、ソノラントエネルギの変動の輪郭２４６の表示領域に示された太い縦線とは擬似音節の境界を示す。音声データ波形２４０の表示領域に示された細い縦線２３０と、ソノラントエネルギの変動の輪郭２４６および基本周波数の波形輪郭２４４の表示領域に示された細い縦線は擬似音節核の境界を示す。

【００７８】

非信頼性の波形２５２の表示領域に示された縦線は波形の極小値部分（ディップ）であり、そこを中心としてＡＱ指数が計算されている部分が最も信頼性の高い部分である。なおＡＱ指数が計算された期間および値は横棒で示されており、横棒の縦位置が高いほど緊張音に近く、低いほど気息音に近い。

【００７９】

図１１には、図１０の点線のボックス２６２で示される時点での声門気流波形の推定値２７０と、その微分波形２７２と、推定された声門気流波形のスペクトル２７４とが示されている。図１０のボックス２６２に対応する時点ではＡＱ指数２５４は高く、すなわちこの時点の発声は緊張音に近い。図１１に示すとおり、このときの声門気流の波形はのこぎり形に近く、サインウェーブの波形からは遠く異なっている。また、微分波形はするどく変化している。

【００８０】

図１２には、図１０の点線のボックス２６０で示される時点での声門気流波形の推定値２８０と、その微分波形２８２と、推定された声門気流波形のスペクトル２８４とが示されている。図１０のボックス２６０に対応する時点ではＡＱ指数２５４は低く、すなわちこの時点の発声は気息音に近い。図１２に示すとおり、このときの声門気流の波形はきれいなサインカーブに近い。微分波形も緩やか

なものとなっている。

【００８１】

上に述べた装置を用い、前述した音声データを実際に処理して擬似音節中心を抽出し、各擬似音節中心に対してＡＱ指数を算出しする一方、それらの擬似音節中心に対応する音を人間が聞いたときに感ずる感想と、ＡＱ指数との相関を以下のようにして調査した。

【００８２】

上記した装置を用いて抽出された信頼性の中心は２２，０００個であり、その各々について対応する声門気流波形およびＡＱ指数と、もとの音声波形のＲＭＳ（Ｒｏｏｔ　Ｍｅａｎ　Ｓｑｕａｒｅ）エネルギ（ｄＢ）とを算出した。これら信頼性の中心のうち、同一の音節核中に存在しかつ互いのＡＱ指数がほぼ一致しているものをまとめ、さらにそれら信頼性の中心のうち、統合された非信頼性の値が０．２以上のものを棄却することにより、聴覚刺激として使用可能と思われる音節核の数は１５，０００をわずかに超えたものとなった。

【００８３】

このデータセットに対して算出された統計情報に基づき、知覚上の評価を行なうために６０の刺激からなるサブセットを選択した。具体的には、前述した３つの感情を表わすデータベースの各々について、極めて低い、または極めて高い、または各感情に対するＡＱ指数の平均値マイナスその分布の標準偏差（σ）近辺、またはＡＱ指数の平均値プラス標準偏差近辺、の４つのカテゴリのいずれかにＡＱ指数が属するような信頼性の中心を含む音節核を５つずつ選択した。

【００８４】

このようにして選択された６０個の擬似音節核の時間的長さは３２ｍｓｅｃから５６０ｍｓｅｃの範囲であり、その平均は１７１ｍｓｅｃであった。通常の聴覚的能力を有する１１人の被験者が、これら短時間の刺激の各々について聴覚的評価を行なった。被験者は静粛なオフィス環境で、高音質のヘッドフォンを用い、各刺激を必要な回数だけ聞き、各刺激について、それぞれ「気息性」および「強さ」とだけ説明した二つのスケールにしたがい、７段階で採点した。各被験者の採点は各々比例により［０，１］の範囲に正規化され、正規化した点数に基づ

出証特２００３－３０２２８１６

いて、６０個の刺激の各々についての１１人の被験者全ての気息性および強さに関する平均値を算出した。

【００８５】

図１３は、上のようにして調べた気息性と、音響的に測定したＡＱ指数の値とを比較する散布図である。これら６０対の値に対する線形相関係数は０．７７であった。この相関は必ずしも高いものではないが、刺激に対するＡＱの測定値が高くなれば、その刺激に対して感じられる気息性も平均すれば高くなるという明らかな傾向があることを裏付けるものといえる。図１３の散布図上で想定されるベストフィットの直線から最も遠い位置に存在する点のいくつかをより詳細に調べると、誤差の原因として次のようなものが浮かび上がる。すなわち、動的制約が欠如しているために生ずる、５つのフレーム中でのフォルマントの非連続性、５つのフレームに含まれていない音節核の一部において生ずる高い気息性、および５つのフレーム中の母音部分に対して、隣接した鼻音がおよぼす強い影響などである。

【００８６】

さらに、図１３からは、中位から下位のＡＱ指数を有する刺激に対しては、気息性の感じ方が広いことに気づく。これは、気息性が低い刺激に気息性に関する点数をつけることが難しく、むしろ地声または緊張音的な発音という側面から点数付けしたほうがよりよく特徴を表わせるのではないか、という直感的な理解を裏付けるものと思われる。

【００８７】

ここでは図としては示していないが、強さの感じ方を、同じ信頼性の中心において測定されたＲＭＳエネルギと比較するための散布図も作成した。その相関係数は０．８３となり、より高度な重み付けを用いて強さの感じ方を測定しているわけではないにもかかわらず、その関係の強さを裏付けるものとなっている。

【００８８】

以上のように本実施の形態では、音響・韻律的分析と、ケプストラム分析とを組合せて、（ｉ）録音された自然な発声中の擬似音節の信頼性の中心の位置を決定するための、（ｉｉ）参考文献２で提案されたＡＱ指数により定量化された音

源の属性を測定するための、全くオペレータが介在する必要がない方法および装置を実現した。そして、その方法および装置を用いて行なった音声知覚の実験の結果は、擬似音節核中で知覚された気息性と強い相関を持つ、頑健性をもって測定できる値としてのＡＱ指数の重要性を確認するものであった。実際、前述したような誤差源が存在しているにもかかわらず、ＡＱ指数と気息性の知覚との間に見出された相関により、音質パラメータとしてのＡＱ指数をさらに研究する必要があることを確認することができた。

【００８９】

そしてこの方法および装置により、発声単位に対するパラ言語的なラベル付けを行なうことができる可能性が高くなる。そうした発声単位を用い、所望のラベル付けがされた発声単位を用いて音声の連続合成を行なうことにより、緊張音から地声、さらに気息的な発音までの範囲にわたる幅広い発声スタイルを用いたマンマシンインタフェースを実現することが可能となる。

【００９０】

[参考文献]

（１） Sundberg, J. (1987). The science of the singing voice, Northern Illinois University Press, Delcalb, Illinois.

（２） Alku, P. & Vilkman, E. (1996). “Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering” , Speech Comm., 18(2), 131-138.

（３） Hermes, D. (1988). “Measurement of pitch by subharmonic summation” , J. Acoust. Soc. Am. 83(1), 257-264.

（４） Mermelstein,
P. (1975). “Automatic segmentation of speech into syllabic units” , J. Acoust.
Soc. Am. 58(4), 880-883.
（５） Lea,
W.A. (1980). “Prosodic aids to speech recognition” , in Lea, W.A. (ed.) , Trends
in Speech Recognition, Prentice-Hall, New Jersey, 166-205.
（６） Broad,
D.J. & Clermont, F. (1989). “Formant estimation by linear transformation of
the LPC cepstrum” , J. Acoust. Soc. Am. 86 (5), 2013-2017.
（７） Mokhtari,
P., Iida, A. & Campbell, N. (2001). “Some articulatory correlates of emotion variability in speech : a
preliminary study on spoken Japanese vowels” , Proc. Int. Conf. on Speech
Process., Taejon, Korea, 431-436.
（８） Peterson,
G.E., & Shoup, J.E. (1966). “A
physiological theory of phonetics” , J. Speech Hear. Res. 9, 5-67.
（９） Iida,
A., Campbell, N., Iga, S., Higuchi, F. & Yasumura, M. (1998). “Acoustic nature and perceptual testing of corpora of emotional speech” , Proc. 5th Int.
Conf. on Spoken Lang. Process., 1559-1562.

【図面の簡単な説明】

【図１】 本発明の一実施の形態のプログラムを実行するコンピュータシステムの外観を示す図である。

【図２】　図１に示すコンピュータシステムのブロック図である。

【図３】　本発明の一実施の形態のプログラムの全体構成をブロック図形式で示す図である。

【図４】　音声データの構成を模式的に示す図である。

【図５】　図３に示す音響・韻律分析部９２のブロック図である。

【図６】　図３に示すケプストラム分析部９４のブロック図である。

【図７】　図６に示す規格化および統合部１４４のブロック図である。

【図８】　図３に示すフォルマントの最適化部９８のブロック図である。

【図９】　図３に示すＡＱ指数計算部１００のブロック図である。

【図１０】　本発明の一実施の形態のプログラムによる表示例を示す図である。

【図１１】　音声データのうち、緊張音と判断される一時点での声門気流波形の推定値、声門気流波形の微分の推定値、および推定された声門気流波形のスペクトルを示す図である。

【図１２】　音声データのうち、気息音と判断される一時点での声門気流波形の推定値、声門気流波形の微分の推定値、および推定された声門気流波形のスペクトルを示す図である。

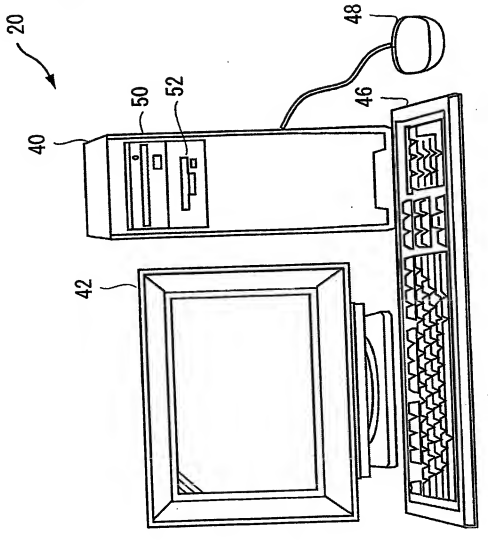
【図１３】　感知された気息性と音響的に測定されたＡＱ指数との間の関連を示す散布図である。

【符号の説明】

２０　コンピュータシステム、８２　音声データ、９０　ＦＦＴ処理部、９２　音響・韻律分析部、９４　ケプストラム分析部、９６　擬似音節中心の抽出部、９８　フォルマントの最適化部、１００　ＡＱ指数計算部、１１０　ピッチ判定部、１１２　ソノラントエネルギ算出部、１１４　ディップ検出部、１１６　有声・エネルギ判定部、１３０　線形予測分析部、１３２　フォルマント推定部、１３４　Δケプストラム算出部、１３６　ケプストラム再生成部、１３８　フレーム間分散算出部、１４０　対数変換および逆ＤＣＴ部、１４２　ケプストラム距離計算部、１４４　規格化および統合部、１４６　信頼性の中心候補出力部、１８６　距離最小化処理部

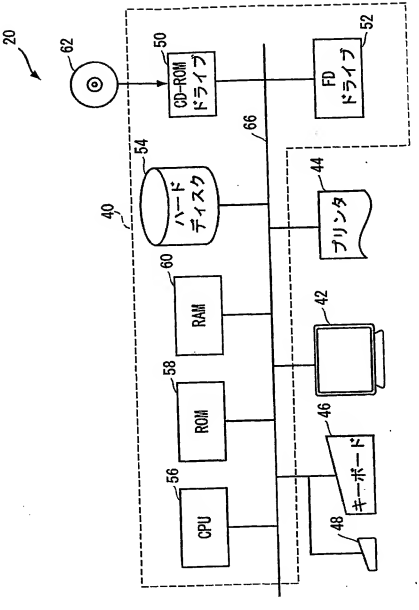
【書類名】　図面

【図１】

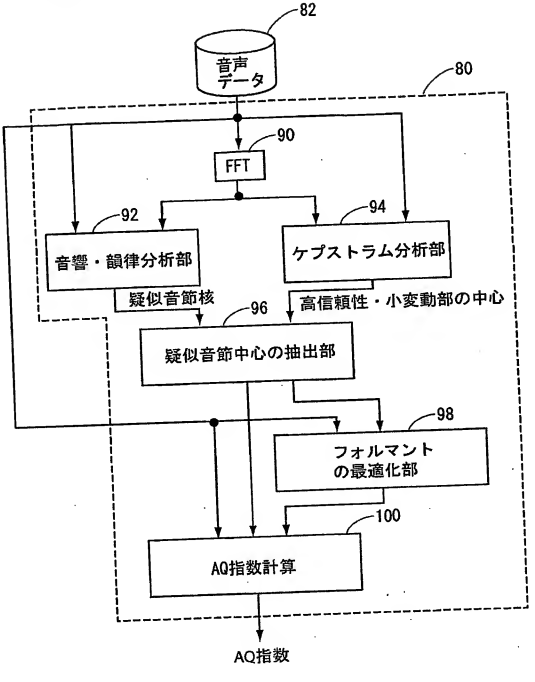


出証特２００３－３０２２８１６

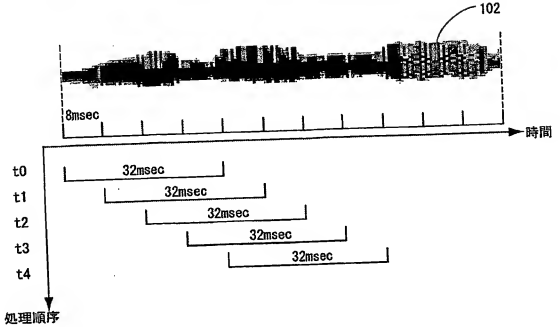
【図２】


20
62
50
CD-ROM
ドライブ
52
FD
ドライブ
66
54
ハード
ディスク
44
プリンタ
40
60
RAM
42
58
ROM
46
キーボード
56
CPU
48

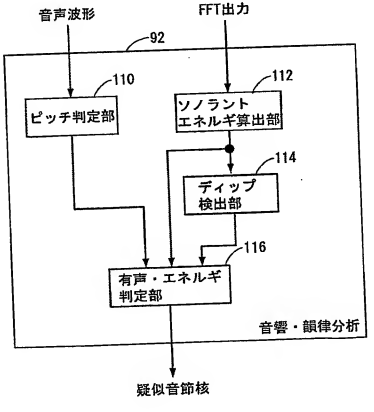
【図３】


82
音声
データ
80
90
FFT
92
94
音響・韻律分析部
ケプストラム分析部
疑似音節核
高信頼性・小変動部の中心
96
疑似音節中心の抽出部
98
フォルマント
の最適化部
100
AQ指数計算
AQ指数

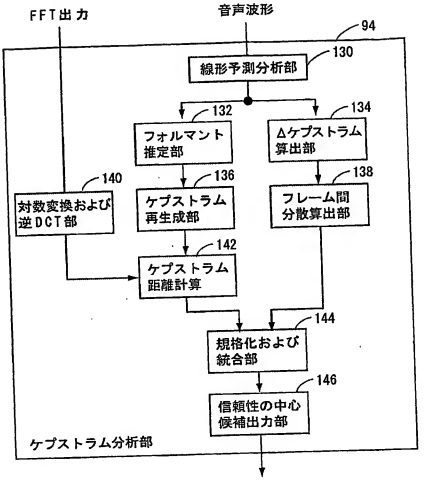
【図４】


102
8msec
時間
t0
t1
t2
t3
t4
32msec
32msec
32msec
32msec
32msec
処理順序

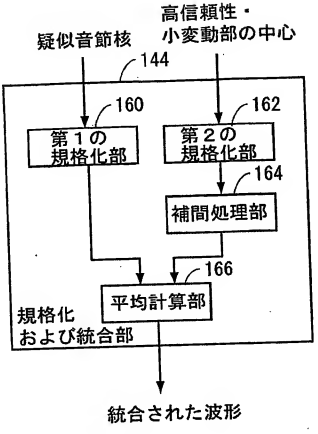
【図５】


音声波形
FFT出力
92
110
ピッチ判定部
112
ソノラント
エネルギ算出部
114
ディップ
検出部
116
有声・エネルギ
判定部
音響・韻律分析
疑似音節核

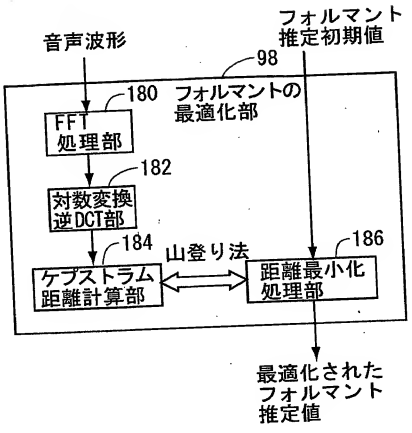
【図６】


FFT出力
音声波形
94
線形予測分析部
130
フォルマント推定部
132
Δケプストラム算出部
134
対数変換および逆DCT部
140
ケプストラム再生成部
136
フレーム間分散算出部
138
ケプストラム距離計算
142
規格化および統合部
144
信頼性の中心候補出力部
146
ケプストラム分析部

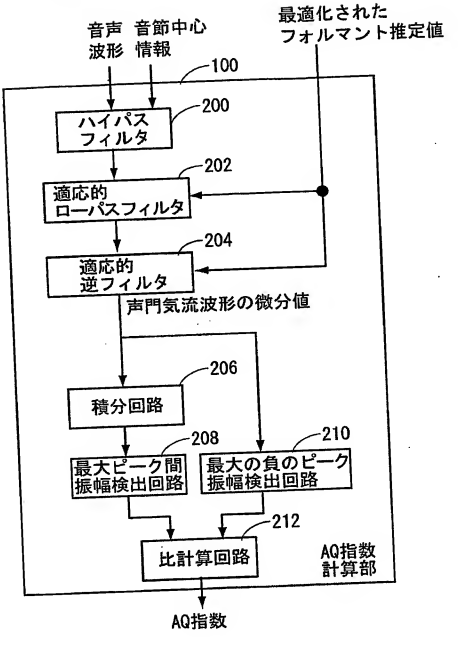
【図７】


疑似音節核
高信頼性・
小変動部の中心
144
160
162
第１の
規格化部
第２の
規格化部
164
補間処理部
166
平均計算部
規格化
および統合部
統合された波形

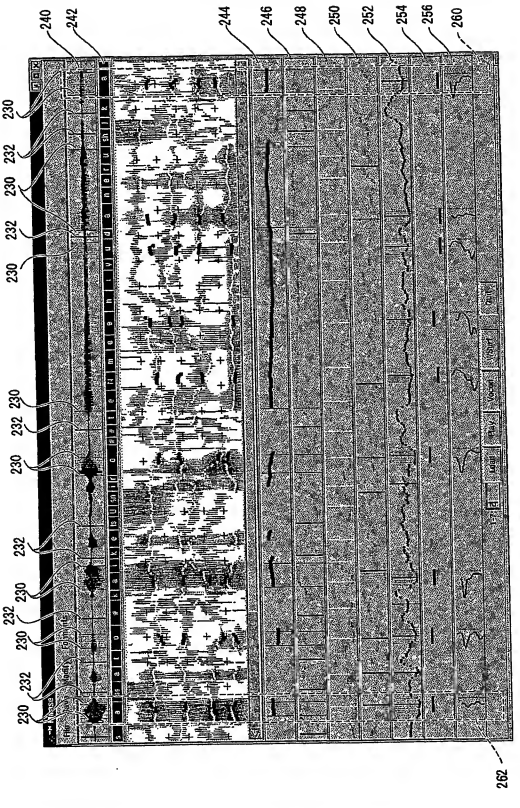
【図８】


音声波形
フォルマント
推定初期値
98
180
フォルマントの
最適化部
FFT
処理部
182
対数変換
逆DCT部
184
山登り法
186
ケプストラム
距離計算部
距離最小化
処理部
最適化された
フォルマント
推定値

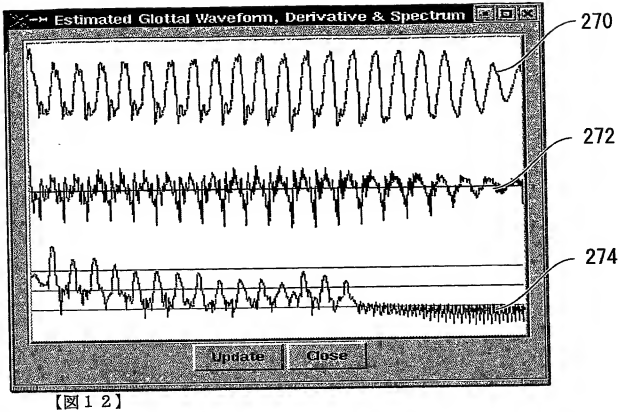
【図９】


音声 音節中心
波形 情報
最適化された
フォルマント推定値
100
ハイパス
フィルタ
200
適応的
ローパスフィルタ
202
適応的
逆フィルタ
204
声門気流波形の微分値
積分回路
206
最大ピーク間
振幅検出回路
208
最大の負のピーク
振幅検出回路
210
比計算回路
212
AQ指数
計算部
AQ指数

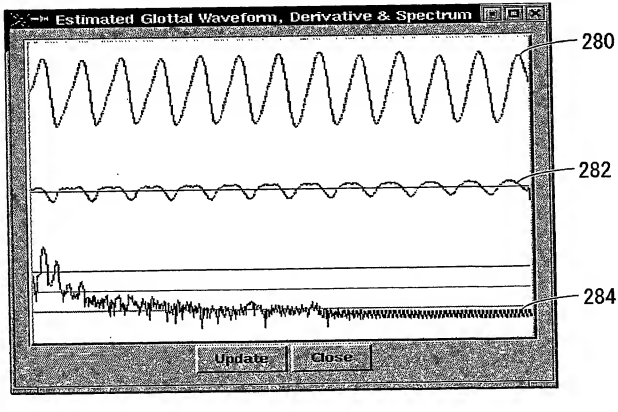
【図１０】



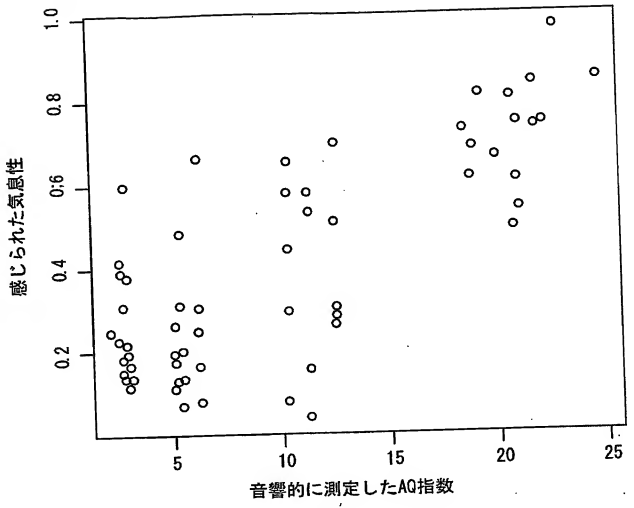
【図１１】



【図１２】



【図１３】



【書類名】　要約書

【要約】

【課題】　音声波形の特徴を高い信頼性で示す部分を決定できるようにする。

【解決手段】　この装置は、データから音声波形のうちの所定周波数領域のエネルギーの時間軸上の分布を算出し、当該分布および音声波形のピッチに基づいて、音声波形の各節のうち、話者によって安定して発生されている領域を抽出する音響・韻律分析部９２と、データから音声波形のスペクトルの時間軸上の分布を算出し、その時間軸上の分布に基づいて、音声波形のうち、その変化が話者により良好に制御されている領域を推定するケプストラム分析部９４と、話者によって安定して発生されている領域として抽出され、かつ話者によってその変化が良好に制御されていると推定された領域を音声波形の高信頼性部分として決定する擬似音節中心の抽出部９６とを含む。

【選択図】　図３

出証特２００３－３０２２８１６

【書類名】　手続補正書
【整理番号】　0020001
【提出日】　平成15年 1月 9日
【あて先】　特許庁長官殿
【事件の表示】
　【出願番号】　特願2002-141390
【補正をする者】
　【識別番号】　396020800
　【氏名又は名称】　科学技術振興事業団
【補正をする者】
　【識別番号】　393031586
　【氏名又は名称】　株式会社国際電気通信基礎技術研究所
【代理人】
　【識別番号】　100099933
　【弁理士】
　【氏名又は名称】　清水　敏
【手続補正　1】
　【補正対象書類名】　特許願
　【補正対象項目名】　発明者
　【補正方法】　変更
　【補正の内容】
　　【発明者】
　　【住所又は居所】　埼玉県川口市本町４丁目１番８号　科学技術振興事業団内
　　【氏名】　モクタリ　パーハム
　　【発明者】
　　【住所又は居所】　京都府相楽郡精華町光台二丁目２番地２　株式会社国際電気通信基礎技術研究所内
　　【氏名】　キャンベル　ニック

出証特２００３－３０２２８１６

【その他】　　出願時発明者名の誤記を訂正するため
【プルーフの要否】　要

認定・付加情報

特許出願の番号	特願２００２－１４１３９０
受付番号	５０３０００２５７９４
書類名	手続補正書
担当官	塩野　実　　２１５１
作成日	平成１５年　１月１６日

＜認定情報・付加情報＞

【補正をする者】

【識別番号】　396020800

【住所又は居所】　埼玉県川口市本町４丁目１番８号

【氏名又は名称】　科学技術振興事業団

【補正をする者】

【識別番号】　393031586

【住所又は居所】　京都府相楽郡精華町光台二丁目２番地２

【氏名又は名称】　株式会社国際電気通信基礎技術研究所

【代理人】

【識別番号】　100099933

【住所又は居所】　大阪府大阪市北区西天満２丁目３番９号　オーク西天満ビル３階　清水　敏特許事務所

【氏名又は名称】　清水　敏

次頁無

出願人履歴情報

識別番号　〔３９６０２０８００〕

１．変更年月日　１９９８年　２月２４日
［変更理由］　名称変更
住　所　埼玉県川口市本町４丁目１番８号
氏　名　科学技術振興事業団

出証特２００３－３０２２８１６

出願人履歴情報

識別番号　〔３９３０３１５８６〕

１．変更年月日　２０００年　３月２７日
［変更理由］　住所変更
住　所　京都府相楽郡精華町光台二丁目２番地２
氏　名　株式会社国際電気通信基礎技術研究所

出証特２００３－３０２２８１６